# CLOUD DATABASES

GROUP A5

AVERY QUICK, CATHERINE TAN, WILL SHAW, VINEETH NAREDDY

WHAT IS A CLOUD DATABASE?

- Hosted and maintained on a cloud computing platform

- Organize, store, manage data for organizations

- Reduces operational costs (e.g. hardware purchases and maintenance, updates, security)

- Different levels:

  - ○ SaaS (Software as a Service)

  - ○ PaaS (Platform as a Service)

  - ○ IaaS (Infrastructure as a Service)

https://cloud.google.com/learn/what-is-a-cloud-database
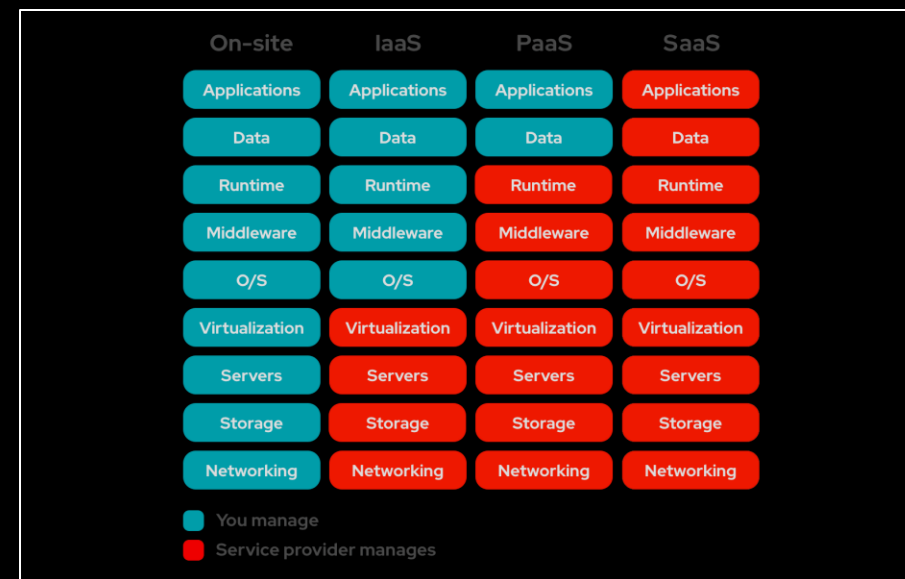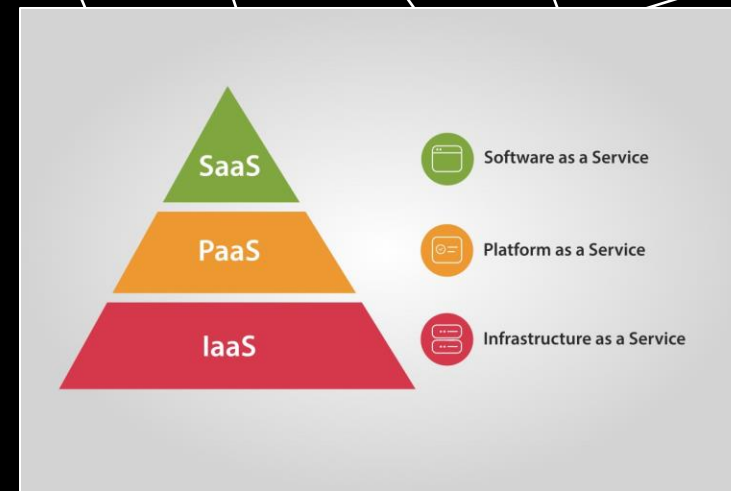
PRODUCT OVERVIEW
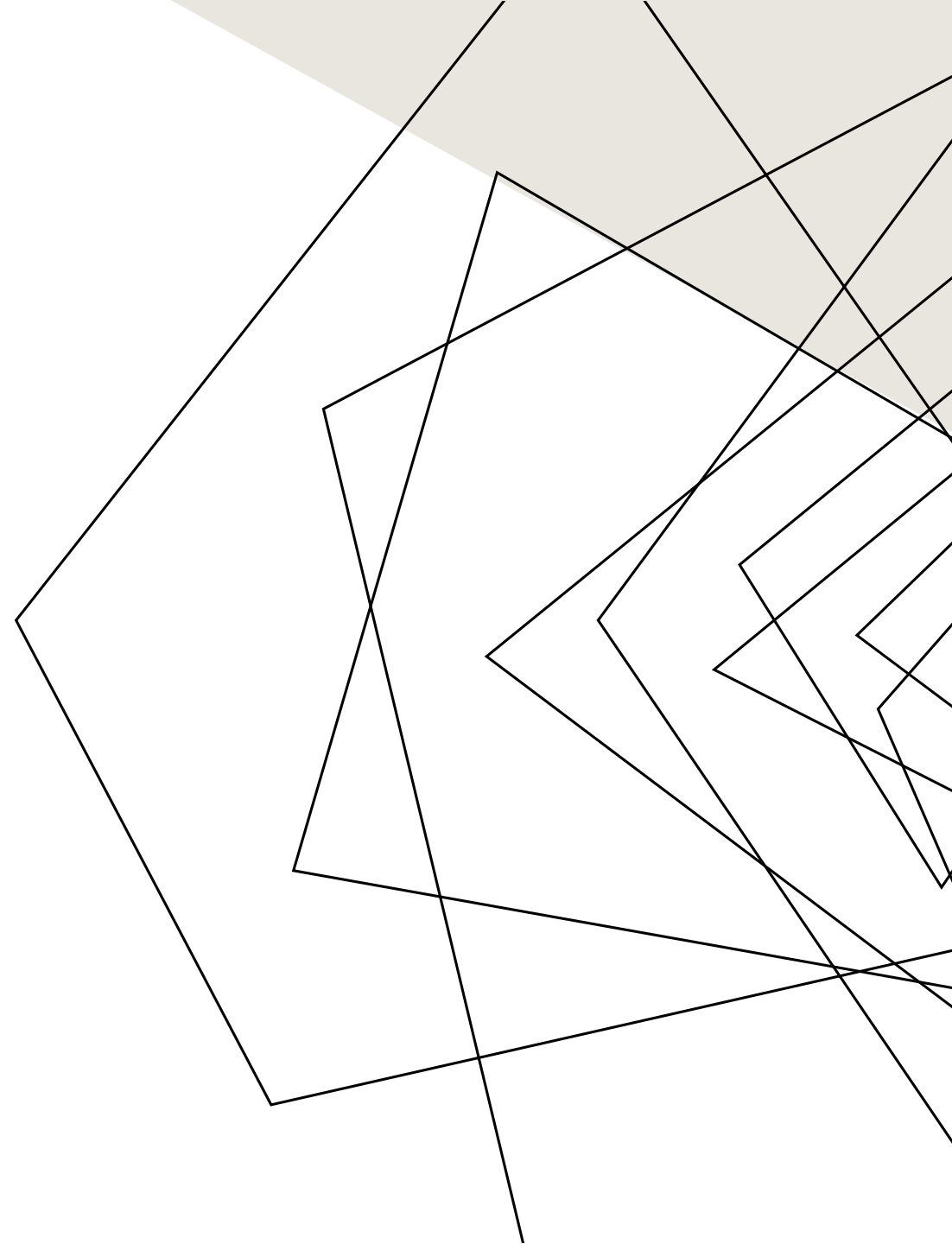
COMPARISON CHART

- **Features and functionalities**
- **Database types**
- **Scalability**
- **Availability**
- **Security**
- **Cost efficiency/Cost model**
- **Use cases**

Feature	Amazon Aurora (AWS)	Oracle	Snowflake
Database Type	Relational (MySQL/PostGreSQL)	Relational (SQL)	Relational (SQL)
Scalability	Seamless and automatically adjusts based on demand	Works well for predefined workloads	Ideal for fluctuating workloads
Availability	High	High	High
Security	Custom encryption, threat detection	Autonomous patching, SQL firewalls, advanced data encryption	Robust encryption
Cost efficiency	Pay as you go, Most cost efficient for small to medium applications	Pay as you go, Less competitive for small to medium applications	Pay as you go, higher costs for high frequency transactional workloads
Best Use Cases	Traditional Relational DB workloads	Transactional, analytical, OLTP workloads	Complex analytical tasks

PRODUCT 1
AMAZON AURORA (AWS)

High performance and scalability — Open all

Up to 5x the throughput of MySQL and 3x the throughput of PostgreSQL +

Serverless configuration +

Automated horizontal scaling +

Features and functionalities
- **Database type: Relational, MySQL & PostgreSQL**
- **Scalability: High, horizontal**
  - **Throughput: 5x on MySQL, 3x on PostgreSQL**
- **Availability: High**
- **Security: High, network isolation, encryption, advance auditing**
- **Cost efficiency/Cost model: pay per use**
- **Use case examples:**
  - **High-traffic Web Apps: large e-commerce companies (Shopify, Walmart, Instacart)**
  - **SaaS Apps: medium to large SaaS providers (Notion, Slack)**
  - **Gaming Applications: large gaming companies (EA, Riot, Epic)**
  - **Financial & Banking Apps: large finance companies, startups (Chime, Robinhood)**

High availability and durability — Open all

Instance monitoring and repair +

Multi-AZ deployments with Amazon Aurora Replicas +

Amazon Aurora Global Database +

Fault-tolerant and self-healing storage +

Automatic, continuous, incremental backups and point-in-time restore +

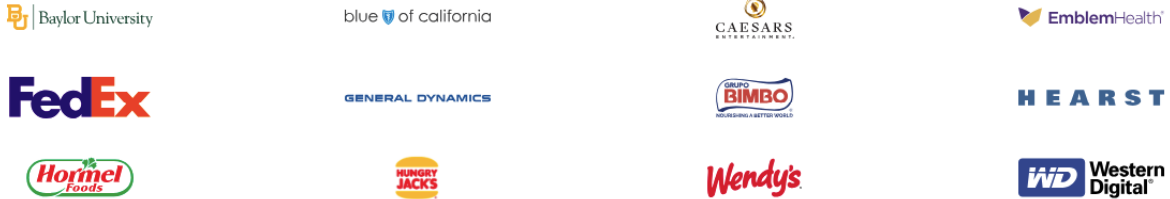Database snapshots +

Backtrack for Aurora MySQL +

https://aws.amazon.com/rds/aurora/features/

https://pganalyze.com/customers/how-notion-runs-postgres-at-scale-on-amazon-rds

https://aws.amazon.com/blogs/database/level-up-your-games-with-amazon-aurora/

PRODUCT 2 ORACLE

Oracle Applications customer successes

See all customer stories

See how leading companies around the world trust Oracle to run their mission-critical business functions.

Baylor University · blue of california · CAESARS · EmblemHealth
FedEx · GENERAL DYNAMICS · GRUPO BIMBO · HEARST
Hormel Foods · HUNGRY JACK'S · Wendy's · Western Digital

Features and functionalities

- **Database type: Relational, SQL**
- **Scalability: High, horizontal & vertical, RAC**
- **Availability: Very High, supports active data guard, multi-region failover**
- **Security: High, encryption, advance auditing, access control**
- **Cost efficiency/Cost model: traditional or pay as you go**
- **Use case examples:**
  - **Enterprise Resource Planning: large enterprises (Fortune 500, Discover, Oracle Netsuite)**
  - **Financial & Banking Apps: large finance companies (JPMorgan & Chase, Bank of America)**
  - **High Volume Transaction Systems: large scale e-commerce (AT&T, Verizon)**
  - **Business Intelligence & Data Warehousing: large corps with massive data analytics (IBM, Accenture)**

Oracle Fusion Cloud Enterprise Resource Planning

Financial Management
Project Management
Procurement
Risk Management and Compliance
Enterprise Performance Management (EPM)
ERP Analytics

Get a clear financial picture to make better decisions

Oracle Financials gives you a complete view of your financial position and results so you can respond quickly to an ever-changing business environment.

See Financials product details

Take a Financials tour

Features
- Accounting hub
- Reporting and analytics
- Payables and Assets
- Revenue management
- Receivables
- Collections
- Expense management
- Joint venture management
- U.S. Federal Financials

https://www.oracle.com/applications/

https://www.oracle.com/erp-0/

PRODUCT 3 SNOWFLAKE

Features and functionalities
- **Database type: Relational, SQL, semi-structured data**
- **Scalability: High, automatic, elastic**
- **Availability: Very High, multi-region failover, reapplication**
- **Security: High, encryption, access control, enterprise grade**
- **Cost efficiency/Cost model: pay as you go**
- **Use case examples:**
  - **Big Data & Analytics: large enterprises with real-time analytics (Netflix, Capital One, Adobe)**
  - **Retail & E-commerce: large retailers (Walmart, Instacart)**
  - **Financial Services & Risk Analysis: banks, companies handling risk modeling (JPMorgan, Stripe)**

https://ieeexplore.ieee.org/abstract/document/9769769

https://docs.snowflake.com/en/user-guide/intro-key-concepts

https://www.snowflake.com/en/pricing-options/

SAMPLE APPLICATIONS

Amazon Web Services

- **90 million active accounts**

- **Wanted dedicated DB for "Story" feature to display interactions on home feed**

- **Required scalability with consistent, high performance and data encryption**

- **Adopted Amazon DynamoDB**

- **Was using on-premises DBs in 13 data centers**

- **Planning and maintenance grew time consuming**

- **Needed system that could support large-scale global game launches, minimize latency, increase resiliency, & reduce downtime**

- **Adopted Amazon RDB then Amazon Aurora**

Venmo Case Study
Riot Games Case Study

MARKET ANALYSIS

Cloud-Based Database Market valued at $22.3 billion in 2023
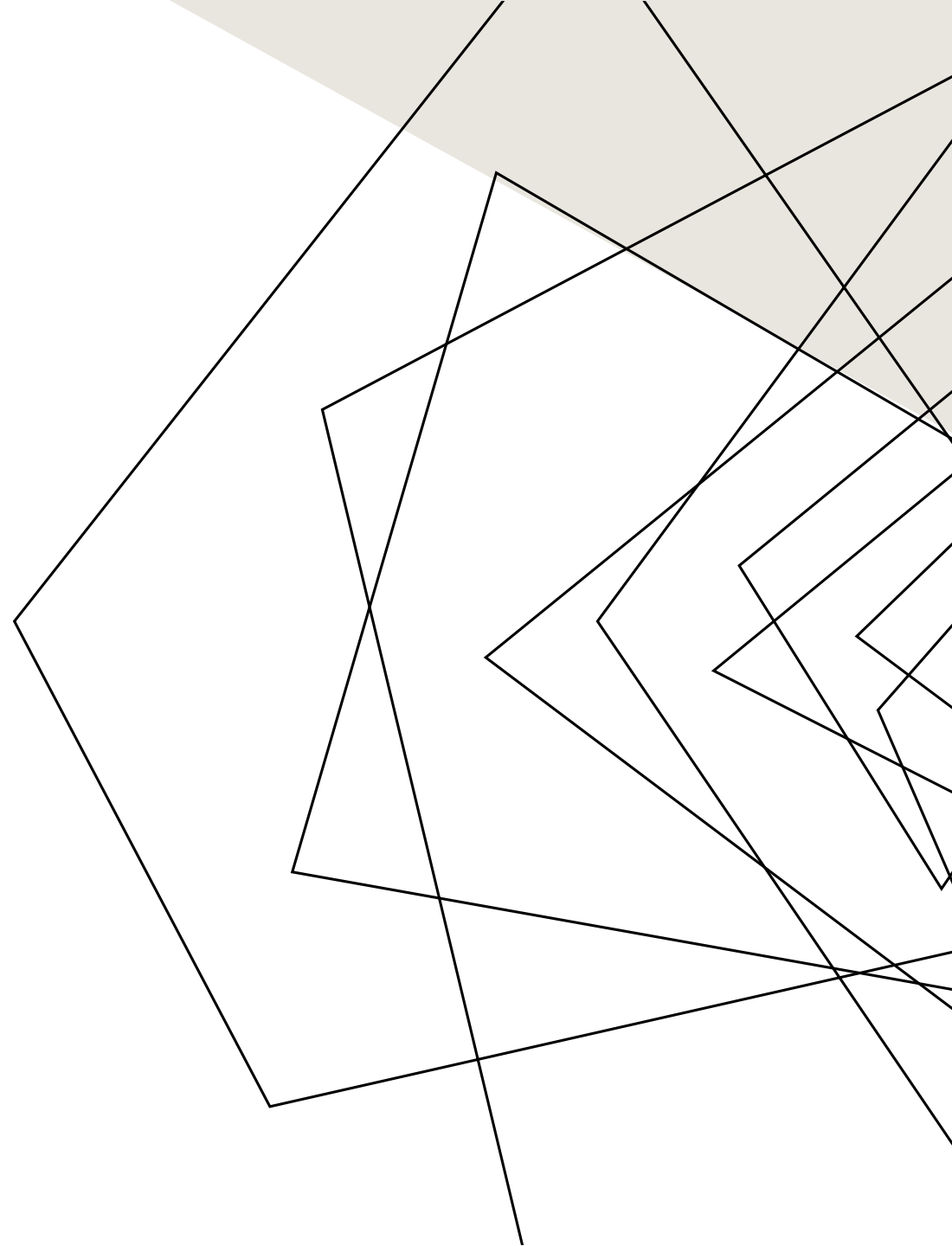
  o Est. to reach $123.4 billion by 2032

Q4 Revenue/Annual Run Rate	Sales Growth Year Over Year	Cloud Market Share For Q4 2024
AWS: $28.8B/$115.2B	AWS: 19%	AWS: 30%
Microsoft: $25.5B/$102B	Microsoft: 19%	Microsoft: 21%
Google Cloud: $12B/$48B	Google Cloud: 30%	Google Cloud: 12%

CRN 2024 Q4 Comparison

TECHNICAL DETAILS

AURORA DB CLUSTER OVERVIEW

What is it?

- **Cloud DB storage volume located in one AWS region**
- **Utilizes Aurora DB instance(s) and a cluster volume to manage SQL data for all instances**
- **Cluster volume allows us to store more data than a single DB instance and stores in a unified location**

Key Components

- **Cluster Volume – manages data from all DB instances and allows us to store more than a single DB instance**
- **Primary (writer) DB Instance – can read from and write to the cluster volume**
- **Reader instance – can only read from the cluster volume used and helps get data from storage cluster in case the primary instance isn't available**
- **S3 Bucket – Used as backup of cluster volume, but allows for additional features like versioning**

11
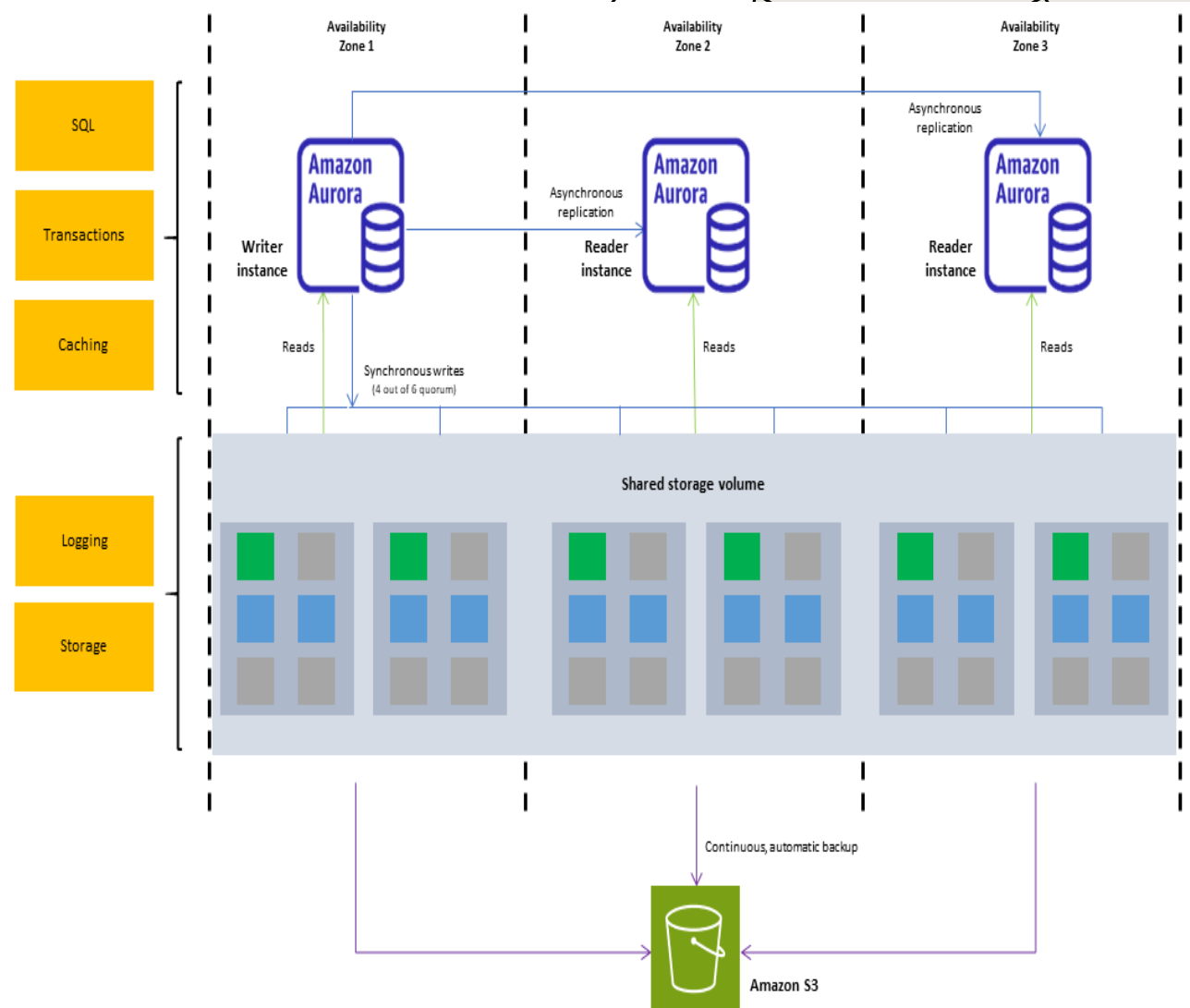
AURORA DB CLUSTER INSIGHTS

Key Technical Differentiator

- **Availability: Use of reader instance ensures that data is accessible**
- **Backups: Use of S3 bucket versioning can help with storing backups of the storage cluster**
- **Optimized I/O Performance: Use of Aurora DB instances help with read/write operations of storage cluster**

Use Cases

- **Sumzap – Japanese smartphone gaming company**
  - o **Uses several Aurora DB clusters with EC2 instances in same availability zone to help with low latency and dealing with peak events**
- **Emma – automated email marketing platform**
  - o **Aurora DB cluster is highly scalable with databases storing 300 million transactions per day**

Summary

- **Aurora DB Clusters are great to use to ensure scalability, availability, etc when working with large amounts of data**

12

SNOWFLAKE OVERVIEW

What is it?

- **Data warehousing platform that optimizes data storage, query processing, and cloud services**
- **Accessible through different cloud platforms (AWS, Google Cloud, Microsoft Azure) through a VPC**

Key Components

- **Database storage – When data is uploaded, Snowflake converts it into micro-partitions between 50 MB to 500 MB and organized into columnar format**
  - **Benefits: columnar store and small size allows for efficient scanning and data manipulation for query processing**
- **Query Processing – Utilizes virtual warehouses to run many queries concurrently**
  - **Virtual warehouses – Cluster of compute resources that use parallel computing techniques to process queries**
- **Cloud services – Services that can help manage the Snowflake platform and processing of user queries**
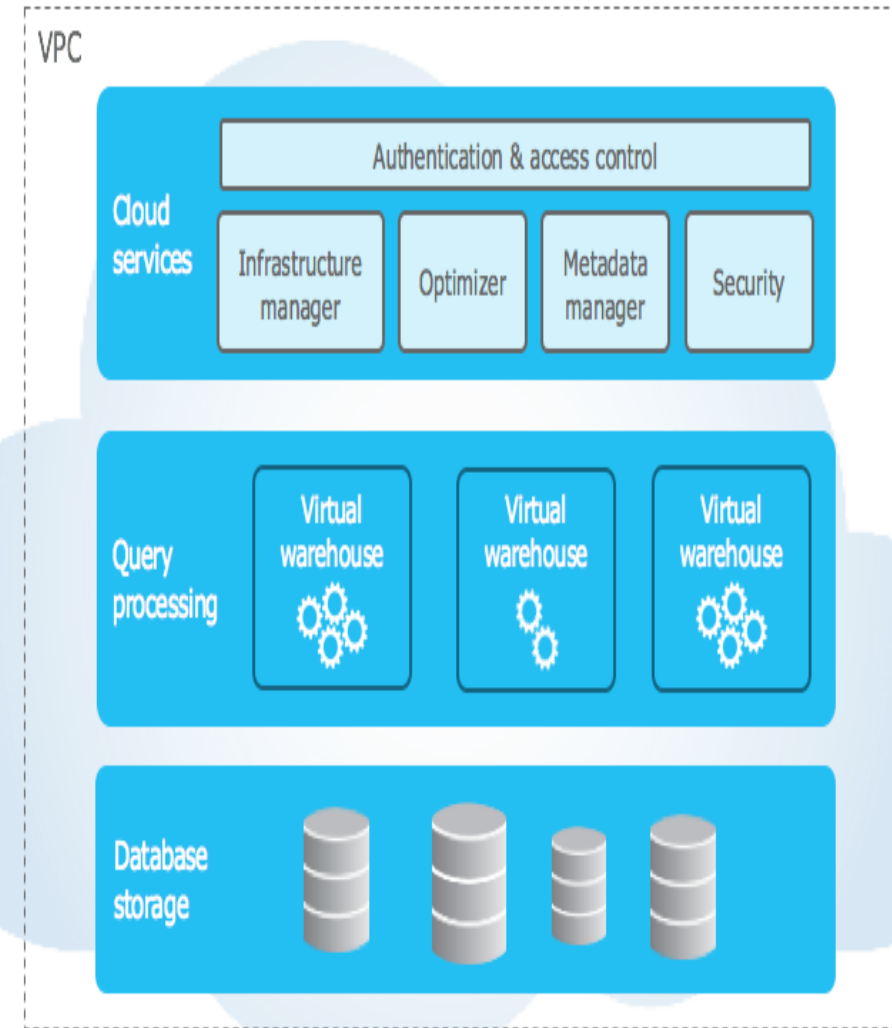
SNOWFLAKE INSIGHTS

Key Technical Differentiator
- **Optimized storage and query processing**
  - Separates both these processes and converts uploaded data into columnar format that can be used by virtual warehouses for processing queries
  - Virtual warehouses allow for processing of many concurrent queries
- **Accessibility:**
  - Can be integrated on all 3 major cloud platforms (AWS, Google Cloud, Microsoft Azure)
  - Easy to upload data stored on cloud to Snowflake
  - Use of VPC ensures that it remains separate of rest of your cloud applications

Use Cases
- **Netflix – largest streaming service in the world**
  - Uses Snowflake to analyze advertising data to figure out what ads to display to a user
- **JP Morgan - one of the largest financial services companies in the world**
  - Uses Snowflake Financial Services Data Cloud to help create a new product (Securities Services Data Mesh) to help investors access investment data
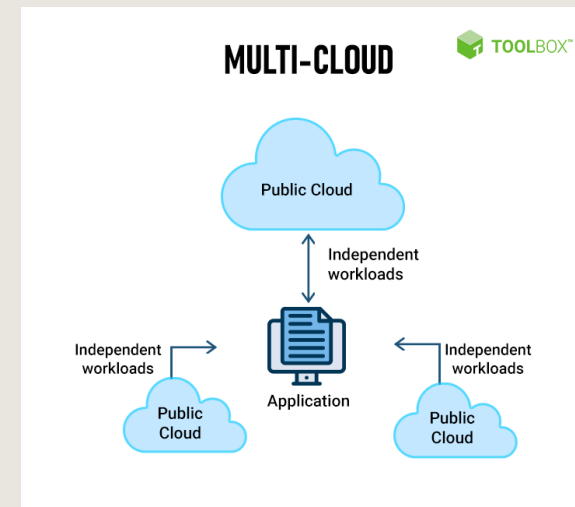   - using Snowflake's analytical and data security features

Summary
- **Snowflake is used by clients because of their data analytical features and develop new features using data (personalized ad recommender)**
- **Great at working for large amounts of data because of accessibility to other cloud services and optimized storage and query processing**
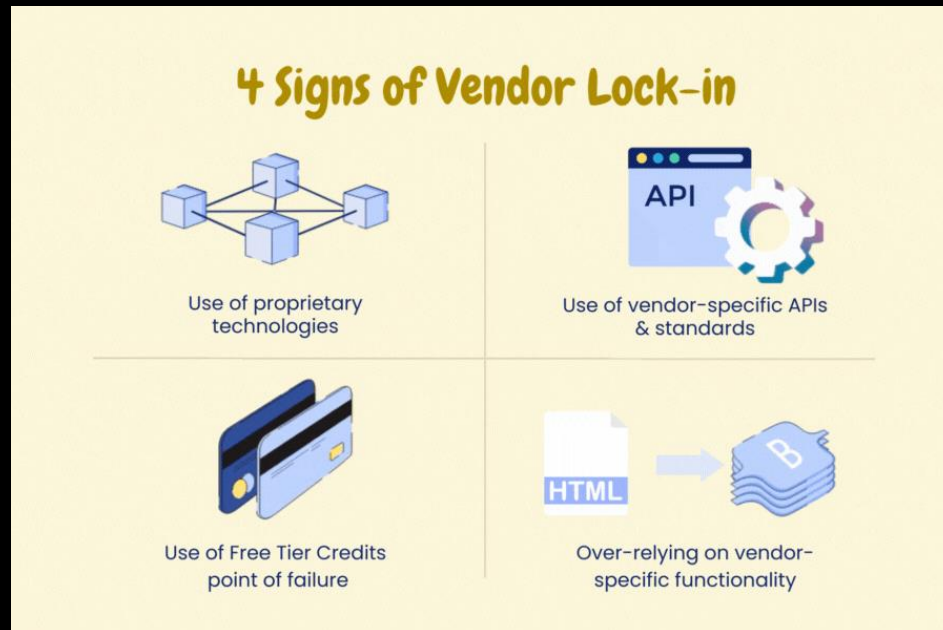
14

FUTURE TRENDS

- AI & Machine Learning Integration
  - Already seeing this with Oracle Autonomous Database
  - Autonomous data management

- Serverless Databases
  - Amazon Aurora Serverless, Google Firestore

- Blockchain
  - decentralized and immutable data

- Multi-Cloud Optimization
  - Unified Database Management Tools

https://www.researchgate.net/profile/Naresh-Kumar-Miryala/publication/386275409_Emerging_Trends_and_Challenges_in_Modern_Database_Technologies_A_Comprehensive_Analysis/links/674b41dd359dcb4d9d46352f/Emerging-Trends-and-Challenges-in-Modern-Database-Technologies-A-Comprehensive-Analysis.pdf

CHALLENGES/ONGOING RESEARCH

4 Signs of Vendor Lock-in

Use of proprietary technologies

Use of vendor-specific APIs & standards

Use of Free Tier Credits point of failure

Over-relying on vendor-specific functionality

https://www.ioriver.io/blog/how-to-avoid-vendor-lock-in

- Data synchronization
  - o Different between cloud providers
  - o Eventual Consistency vs. Strong Consistency
- Security
  - o Wiz – $23 billion
- Vendor Lock-In
  - o Databases can become very entrenched

https://www.calcalistech.com/ctechnews/article/hjpwti2dr